IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPEAL BRIEF – 37 C.F.R § 1.192

U.S. Patent Application 10/605,208 entitled:

"AUTOMATIC QUERY ROUTING AND RANK CONFIGURATION FOR SEARCH QUERIES IN AN INFORMATION RETRIEVAL SYSTEM"

Real Party in Interest: International Business Machines Corporation

Related Appeals and Interferences:

None

Status of Claims:

Claims 1-19 are pending.

Claims 1-19 stand rejected under 35 U.S.C. § 102(B) as being anticipated by Christianson et al. (US 6,085,186).

Claims 1-19 are hereby appealed.

Status of Amendments:

No after-final amendments were submitted after the final amendment of 05/25/2007.

Summary of Claimed Subject Matter:

(NOTE: All citations are made from the corresponding pre-grant publication **US 2005/0060290 A1**, which recites the original specification, including the figures.)

As per independent claim 1, the present invention provides a method for identifying documents most relevant to a query from a collection of documents that is organized based on a set of indices, said method comprising the steps of: (a) determining a query class for a received query based on statistical information regarding query terms of said received query and lexical affinities associated with permutations of said query terms, said query class associated with a routing function and a ranking function, said routing function capable of determining subsets of

the collection that most likely include the most relevant documents, and said ranking function capable of sorting the documents in terms of relevancy **(see paragraph [0037], figure 2, step 202)**; (b) identifying a set of indices most relevant to said query **(see paragraph [0037], figure 2, step 204)**; (c) identifying a set of documents related to said query based on said determined indices, said identification performed via passing said ranking function associated with said determined query class along with said query to each search engine that manages a determined index from a collection of relevant indices **(see paragraph [0037], figure 2, step 206)**; (d) collecting results ranked based upon said ranking function and merging and sorting said collected results by relevancy **(see paragraph [0037], figure 2, step 208)**; and (e) returning a subset of the highest ranked documents as the documents most relevant to the query **(see paragraph [0037], figure 2, step 208)**.

As per independent claim 8, the present invention provides an article of manufacture comprising a computer user medium having computer readable code embodied therein which identifies documents most relevant to a query from a collection of documents that is organized based on a set of indices, said medium comprising: (a) computer readable program code determining a query class for a received query based on statistical information regarding query terms of said received query and lexical affinities associated with permutations of said query terms, said query class associated with a routing function and a ranking function, said routing function capable of determining subsets of the collection that most likely include the most relevant documents, and said ranking function capable of sorting the documents in terms of relevancy **(see paragraph [0037], paragraph [0054], and figure 2, step 202)**; (b) computer

readable program code determining indices most relevant to said query **(see paragraph [0037], paragraph [0054], and figure 2, step 204)**; (c) computer readable program code identifying a set of documents related to said query based on said determined indices, said identification performed via passing said ranking function associated with said determined query class along with said query to each search engine that manages a determined index from a collection of relevant indices **(see paragraph [0037], paragraph [0054], and figure 2, step 206)**; (d) computer readable program code collecting results ranked based upon said ranking function and merging and sorting said collected results by relevancy; and (e) computer readable program code returning a subset of the highest ranked documents as the documents most relevant to the query **(see paragraph [0037], paragraph [0054], and figure 2, step 208)**.

As per independent claim 12, the present invention provides a method for retrieving information comprising the steps of: (a) receiving a query **(see paragraph [0030] through [0036], figure 1A, step 102)**; (b) parsing said query and generating a set of query terms **(see paragraph [0030] through [0036], figure 1A, step 104)**; (c) identifying statistical information regarding each of said query terms and different permutations of query terms **(see paragraph [0030] through [0036], figure 1A, step 108)**; (d) identifying lexical affinities associated with said permutations of query terms **(see paragraph [0030] through [0036], figure 1A, step 110)**; (e) classifying said query into a query category based upon results of steps c and d **(see paragraph [0030] through [0036], figure 1A, step 114)**; (f) identifying a set of ranking parameters associated with said query category **(see paragraph [0030] through [0036], figure 1B, step 116)**; (g) identifying routing information associated with said query category **(see**

paragraph [0030] through [0036], figure 1B, step 118); (h) issuing a query to a search engine by applying said identified ranking parameters and said identified routing information **(see paragraph [0030] through [0036], figure 1B, step 120)**; and (i) receiving and rendering search results from said search engine **(see paragraph [0030] through [0036], figure 1B, step 122)**.

As per independent claim 17, the present invention provides an article of manufacture comprising a computer storage medium having computer readable code embodied therein for retrieving information comprising the steps of: (a) computer readable program code receiving a query **(see paragraph [0030] through [0036], paragraph [0054], figure 1A, step 102)**; (b) computer readable program code parsing said query and generating a set of query terms **(see paragraph [0030] through [0036], paragraph [0054], figure 1A, step 104)**; (c) computer readable program code identifying statistical information regarding each of said query terms and different permutations of query terms **(see paragraph [0030] through [0036], paragraph [0054], figure 1A, step 108)**; (d) computer readable program code identifying lexical affinities associated with said permutations of query terms **(see paragraph [0030] through [0036], paragraph [0054], figure 1A, step 114)**; (e) computer readable program code classifying said query into a query category based upon results of steps c and d **(see paragraph [0030] through [0036], paragraph [0054], figure 1A, step 114)**; (f) computer readable program code identifying a set of ranking parameters associated with said query category **(see paragraph [0030] through [0036], paragraph [0054], figure 1B, step 116)**; (g) computer readable program code identifying routing information associated with said query category **(see paragraph [0030] through [0036], paragraph [0054], figure 1B, step 118)**; (h) computer

readable program code issuing a query to a search engine by applying said identified ranking parameters and said identified routing information **(see paragraph [0030] through [0036], paragraph [0054], figure 1B, step 120)**; and (i) computer readable program code receiving and rendering search results from said search engine **(see paragraph [0030] through [0036], paragraph [0054], figure 1B, step 122)**.

Grounds of Rejection to be Reviewed on Appeal:

Claim 1-19 are pending. Claims 1-19 stand rejected under 35 U.S.C. §102(b) as being anticipated by Christianson et al. (US 6,085,186), hereafter Christianson. Claims 1-19 are hereby appealed. Was a proper rejection made under 35 U.S.C. §102(b) using existing USPTO guidelines?

ARGUMENT:

Claim 1-19 are pending. Claims 1-19 stand rejected under 35 U.S.C. § 102(b) as being anticipated by Christianson et al. (US 6,085,186), hereafter Christianson. Claims 1-19 are hereby appealed. Was a proper rejection made under 35 U.S.C. § 102(b) using existing USPTO guidelines?

REJECTIONS UNDER 35 U.S.C. § 102(b)

To be properly rejected under 35 U.S.C § 102(b), **each and every claim element** must be shown in a single reference (i.e., in this case, the Christianson reference). Applicant respectfully disagrees with the Examiner that the claims are taught by the cited art. The Manual for Patenting Examining Procedure (MPEP) § 2131 clearly sets forth the standard for rejecting a claim under 35 U.S.C. § 102(e). "A claim is anticipated only if each and every element as set forth in the claim is found, either expressly or inherently described, in a single prior art reference." (MPEP § 2131, quoting Verdegaal Bros. v. Union Oil Co. of California 2 USPQ2d 1051, 1053 (Fed Cir. 1987)). In this case, the cited art (i.e., Christianson) fails to teach the claimed invention as required by the MPEP.

Christianson teaches a method for assisting a user to query for information available from information sources attached to a network, wherein the method comprising the steps of: selecting the one or more information sources most relevant to a user query, formatting the user query for each relevant information source according to a description of each relevant information source written in a wrapper description language, and transmitting the formatted query to each of the relevant information sources, extracting data fields relevant to the user query

from responses returned from the relevant information sources, and presenting the relevant data fields to the user.

Applicants' independent claim 1, by stark contrast, provides a method for identifying documents most relevant to a query from a collection of documents that is organized based on a set of indices, said method comprising the steps of: (a) determining a query class for a received query **based on statistical information** regarding query terms of said received query **and lexical affinities associated with permutations of said query terms**, **said query class associated with a routing function and a ranking function**, **said routing function capable of determining subsets of the collection that most likely include the most relevant documents**, and **said ranking function capable of sorting the documents in terms of relevancy**; (b) identifying a set of indices most relevant to said query; (c) identifying a set of documents related to said query based on said determined indices, **said identification performed via passing said ranking function associated with said determined query class along with said query to each search engine that manages a determined index from a collection of relevant indices**; (d) collecting results ranked based upon said ranking function and merging and sorting said collected results by relevancy; and (e) returning a subset of the highest ranked documents as the documents most relevant to the query.

Similarly, Applicants' independent claim 8 provides for an article of manufacture implementing the method of claim 1.

The Examiner, on page 2 of the Office Action of 05/27/2006, asserts that column 7, line 57 through column 8, line 20, column 9, lines 2-18, and column 14, lines 50-65 teach Claim 1's (and claim 8's) feature of "**determining a query class for a received query based on *statistical information* regarding query terms of said received query and *lexical affinities* associated with *permutations of said query terms***".

Christianson's column 7, line 57 through column 8, line 20 merely teaches a "**query router**". By Christianson's own admission in the Examiner's own citation, the query router merely "**determines the relevance of each information source to the given query and returns N more relevant sources**" (see column 8, lines 1-3). Christianson further clarifies in column 8, lines 11-20 that the functionality of the query router as a module that "**calculates a numerical relevance rank for each information source**" wherein this calculation is based on "**conceptual classes**" and further clarifies that "**each information source is tagged in advance with the conceptual classes for which it is relevant**". Christianson yet again clarifies that **each query is mapped (e.g., using hashing functions) to a conceptual class relevant to it and finds information sources with conceptual classes shared by the query**.

It should be emphasized that although Christianson uses similar terminologies, such as conceptual classes, it is evident from the above-citation that Christianson's "conceptual classes" does not teach or suggest the "query class" of Applicants' claim 1 and 8. By Christianson's own admission, **the mapping of a query to "conceptual classes" is done via a "hash function", and NOT based on "statistical information regarding query terms" AND "lexical affinities**

associated with permutations of said query terms".

For further support, Applicants direct the attention of the Examiner to column 4, lines 31-36 (reproduced below), which provide more detail regarding the conceptual classes.

> "**Groups of sources 7 having similar sorts of information are grouped into conceptual classes called information domains**. For example, one domain can be that of electronic stores for a particular product; another domain might include Internet indexes containing information on the keyword content of various World Wide Web ("WWW") pages." (*emphasis added*).

The above-citation clearly indicates that the "conceptual classes" of Christianson are merely a "**grouping of sources [7] having similar sorts of information**" and **CANNOT** be equated to Applicants' query class which, for a received query, is determined based on **statistical information** regarding query terms **AND** ***lexical affinities*** associated with ***permutations of the query terms***.

Applicants, therefore, respectfully assert that **Christianson reference in its entirety fails to teach or suggest determining such "conceptual classes" based on either "statistical information regarding query terms" OR "*lexical affinities* associated with *permutations* of said query terms"**. Independent claims 1 and 8 cannot be rejected under 35 U.S.C. §102(b),

when either of these instances (i.e., determining a query class for a received query based on statistical information OR determining a query class for a received query based on lexical affinities associated with permutations of said query terms) is not shown in a single reference, let alone showing both instances (i.e., determining a query class for a received query based on statistical information AND determining a query class for a received query based on lexical affinities associated with permutations of said query terms) in a single reference, as required by independent claims 1 and 8.

Further, the Examiner's citations of column 9, lines 2-18 and column 14, lines 50-65 merely recites the step of **retrieving a page in order to calculate a "relevance estimate"**. However, as above, there is neither an explicit nor an implicit teaching/suggestion regarding the **determination of a "query class" based on statistical information regarding query terms AND lexical affinities associated with *permutations* of the query terms**. In fact, **there is no mention of finding *lexical affinities* associated with *permutations* of query terms**.

Further, column 9, lines 5-11 merely teach the following: "In an exemplary embodiment, **for queries requesting the presence either of all query words or of any query words**, the estimate is determined by scanning the page and counting the number of query words actually present, and then scaling the count so that the presence of all words results in the common maximum relevance value."

Also, column 9, lines 11-15 of Christianson merely teach the following: "**For queries**

requesting the presence of a phrase, the estimate is determined, for example, by subtracting from the common maximum a normalized sum of the square of the distance in the page of each word of the phrase from its successor word in the phrase"

Hence, by Christianson's own admission, the estimate is determined by scanning the page and counting the query words **if the query requests the presence of either or all query words or any query words**, **OR**, **in another option**, the estimate is determined by subtracting from the common maximum a normalized sum of the square of the distance in the page of each word of the phrase from its successor word in the phrase, **in the case of a query requesting the presence of a phrase**.

Hence, it would be erroneous to combine the two options and conclude from Christianson's teaching that a **determination of a "query class" is made based on statistical information regarding query terms AND lexical affinities associated with permutations of the query terms**.

Hence, at least for the reasons set forth above, Applicants respectfully assert that the Examiner has issued an improper rejection with respect to independent claims 1 and 8.

Applicants' independent claim 12 provides a method for retrieving information comprising the steps of: (a) receiving a query; (b) parsing said query and generating a set of query terms; (c) **identifying statistical information regarding each of said query terms and**

different permutations of query terms; (d) **identifying lexical affinities associated with said permutations of query terms**; (e) **classifying said query into a query category based upon results of steps c and d**; (f) **identifying a set of ranking parameters** associated with said query category; (g) **identifying routing information** associated with said query category; (h) issuing a query to a search engine by applying said identified ranking parameters and said identified routing information; and (i) receiving and rendering search results from said search engine.

Applicants' independent claim 17 provides for a computer product implementing the above-method.

The above-mentioned arguments with respect to independent claims 1 and 8 substantially apply to independent claims 12 and 17. As above, Applicants respectfully assert that Christianson reference in its entirety fails to teach or suggest determining such "conceptual classes" based on either "**statistical information regarding query terms**" **OR** "**lexical affinities associated with permutations of said query terms**". With either of these instances not shown in the Christianson reference, Applicants are unsure how the Examiner can assert that the Christianson reference teaches the step of "classifying said query into a query category based upon results of steps c **and** d".

Hence, at least for the above-reasons, Applicants respectfully assert that the Examiner has issued an improper rejection with respect to independent claims 12 and 17.

Furthermore, the above-mentioned arguments with respect to independent claims 1, 8, 12, and 17 substantially apply to dependent claims 2-7, 9-11, 13-16 and 18-19 as they inherit all the features of the claim from which they depend.

SUMMARY

As has been detailed above, none of the references, cited or applied, provide for the specific claimed details of applicants' presently claimed invention, nor renders them obvious. It is believed that this case is in condition for allowance and reconsideration thereof and early issuance is respectfully requested.

As this Appeal Brief has been timely filed within the set period of response, no fee for extension of time is required. However, the Commissioner is hereby authorized to charge any deficiencies in the fees provided, including extension of time, to Deposit Account No. 09-0441.

Respectfully submitted by
Applicant's Representative,

/Ramraj Soundararajan/

Ramraj Soundararajan
Reg. No. 53832

IP AUTHORITY, LLC
4821A Eisenhower Ave
Alexandria, VA 22304
(703) 461-7060

October 29, 2007

Claims Appendix:

1. A method for identifying documents most relevant to a query from a collection of documents that is organized based on a set of indices, said method comprising the steps of:

a) determining a query class for a received query based on statistical information regarding query terms of said received query and lexical affinities associated with permutations of said query terms, said query class associated with a routing function and a ranking function, said routing function capable of determining subsets of the collection that most likely include the most relevant documents, and said ranking function capable of sorting the documents in terms of relevancy;

b) identifying a set of indices most relevant to said query;

c) identifying a set of documents related to said query based on said determined indices, said identification performed via passing said ranking function associated with said determined query class along with said query to each search engine that manages a determined index from a collection of relevant indices;

d) collecting results ranked based upon said ranking function and merging and sorting said collected results by relevancy; and

e) returning a subset of the highest ranked documents as the documents most relevant to the query.

2. The method as per claim 1, wherein said step for determining a query class further comprises the following steps:

a) analyzing user profile data, user search context and history data, log file data, and

index statistics, or other query related external data; and

b) utilizing said analyzed data in determining a query class for said search query.

3. The method as per claim 1, wherein said step for identifying a set of indices further comprises the step of using routing information obtained from applying said routing function associated with said query class to determine said set of indices.

4. The method as per claim 1, wherein said step of returning a subset of the highest ranked documents further comprises the following steps:

a) assigning each search result item a relevancy score; and

b) returning a predetermined subset of results from said search results.

5. The method as per claim 4, wherein said method additionally comprises the step of sorting search results by said relevancy score in decreasing order prior to returning said predetermined subset of results.

6. A method as per claim 1, wherein said method is implemented across networks.

7. A method as per claim 6, wherein said across networks element comprises any of, or a combination of, the following: wide area network (WAN), local area network (LAN), cellular, wireless, or the Internet.

8. An article of manufacture comprising a computer user medium having computer readable code embodied therein which identifies documents most relevant to a query from a collection of documents that is organized based on a set of indices, said medium comprising:

a) computer readable program code determining a query class for a received query based on statistical information regarding query terms of said received query and lexical affinities associated with permutations of said query terms, said query class associated with a routing function and a ranking function, said routing function capable of determining subsets of the collection that most likely include the most relevant documents, and said ranking function capable of sorting the documents in terms of relevancy;

b) computer readable program code determining indices most relevant to said query;

c) computer readable program code identifying a set of documents related to said query based on said determined indices, said identification performed via passing said ranking function associated with said determined query class along with said query to each search engine that manages a determined index from a collection of relevant indices;

d) computer readable program code collecting results ranked based upon said ranking function and merging and sorting said collected results by relevancy; and

e) computer readable program code returning a subset of the highest ranked documents as the documents most relevant to the query.

9. An article of manufacture as per claim 8, wherein said computer readable program code determining a query class further comprises:

a) computer readable program code analyzing user profile data, user search context and

history data, log file data, and index statistics, or other query related external data; and

b) computer readable program code utilizing said analyzed data in determining a query class for said search query.

10. An article of manufacture as per claim 8, wherein said computer readable program code identifying a set of indices further comprises computer readable program code using routing information obtained from applying said routing function associated with said query class to determine said set of indices.

11. An article of manufacture as per claim 8, wherein said computer readable program code returning a subset of the highest ranked documents further comprises:

a) computer readable program code assigning each search result item a normalized score;

b) computer readable program code sorting search results by score in decreasing order of said scores; and

c) computer readable program code returning a predetermined subset of results from said sorted list of search results.

12. A method for retrieving information comprising the steps of:

a) receiving a query;

b) parsing said query and generating a set of query terms;

c) identifying statistical information regarding each of said query terms and different permutations of query terms;

d) identifying lexical affinities associated with said permutations of query terms;

e) classifying said query into a query category based upon results of steps c and d;

f) identifying a set of ranking parameters associated with said query category;

g) identifying routing information associated with said query category;

h) issuing a query to a search engine by applying said identified ranking parameters and said identified routing information; and

i) receiving and rendering search results from said search engine.

13. A method as per claim 12, wherein said step of identifying statistical information additionally comprises the step of analyzing log data.

14. A method as per claim 12, wherein said step of identifying statistical information additionally comprises the step of analyzing user feedback.

15. A method as per claim 12, wherein said method is implemented across networks.

16. A method as per claim 15, wherein said across networks element comprises any of, or a combination of, the following: wide area network (WAN), local area network (LAN), cellular, wireless, or the Internet.

17. An article of manufacture comprising a computer storage medium having computer readable code embodied therein for retrieving information comprising the steps of:

a) computer readable program code receiving a query;

b) computer readable program code parsing said query and generating a set of query terms;

c) computer readable program code identifying statistical information regarding each of said query terms and different permutations of query terms;

d) computer readable program code identifying lexical affinities associated with said permutations of query terms;

e) computer readable program code classifying said query into a query category based upon results of steps c and d;

f) computer readable program code identifying a set of ranking parameters associated with said query category;

g) computer readable program code identifying routing information associated with said query category;

h) computer readable program code issuing a query to a search engine by applying said identified ranking parameters and said identified routing information; and

i) computer readable program code receiving and rendering search results from said search engine.

18. The method of claim 1 further comprising:

performing steps a-d for each of a plurality of query classes; and weighting results from each search engine for each query class according to a degree of probability to which the query is associated with each of the query classes.

19. The method of claim 12 further comprising:

performing steps f-i for each of a plurality of query categories; and weighting results

from each search engine for each query category according to a degree of probability to which the query is associated with each of the query categories.

Evidence Appendix

None

Related Proceedings Appendix

None